

ABSTRACT OF THE DISCLOSURE

An information computational method for classifying multivariate datasets to identify latent (unobservable) properties of members of a sample, which properties are then used for classification. The method comprises a novel combination of statistical and fuzzy logic methods whereby the latent classes of each object are identified according to the formula:

$$f(\bar{Y}_{j_1, \dots, j_K}) \left| \left\{ j_k \in S_{km_k} \right\}_{k=1}^K - G \left[h \left(k, j_k, \left\{ S_{km} \right\}_{m=1}^{M_k} \right) \right] \right|$$

wherein $k \in \{1, \dots, K\}$ indexes the directions of the multidimensional space; $j_k \in \{1, \dots, N_k\}$ identifies an object in direction k ; N_k is the number of objects in principal direction k ; $\bar{Y}_{j_1, \dots, j_K}$ is a vector of one or more observations on a set of objects $\{j_1, \dots, j_K\}$; $m \in \{1, \dots, M_k\}$ indexes latent classes in direction k with M_k being the number of latent classes in direction k ; S_{km} is a latent class m in direction k ; $G[\cdot]$ is a specified univariate or multivariate distribution; $f(\cdot)$ and $g(\cdot)$ are specified functions; and the method calculates the likelihood that each object of interest belongs to each identified latent class. The invention addresses a variety of informatics problems, particularly in the field of biology, and permits a user to make reasonable inferences about underlying cause-effect relationships, such as the underlying biology of gene-expression patterns.